The Substrate Specificity of a Recombinant Cysteine Protease from *Leishmania mexicana*: Application of a Combinatorial Peptide Library Approach

Phaedria M. St. Hilaire,[a] Lira C. Alves,[b] Sanya J. Sanderson,[c] Jeremy C. Mottram,[d] Maria A. Juliano,[b] Luiz Juliano,[b] Graham H. Coombs,[c] and Morten Meldal*[a]

The substrate specificity of CPB2.8ΔCTE, a recombinant cysteine protease from Leishmania mexicana*, was mapped by screening a fluorescence-quenched combinatorial peptide library. Results from library screening indicated a preference for Arg or Lys in the S_3 subsite and for hydrophobic residues, both aliphatic and aromatic, in S_2. The S_1 subsite exhibited a specificity for the basic residues Arg and Lys. Generally, the specificity of the primed subsites was less strict compared with the non-primed side which showed preference for Arg, Lys and Ala in S_1', Arg, Pro and Gly in S_2' and Lys, Arg and Ser in S_4'. By contrast, a strict preference for the basic residues Arg and Lys was found for S_3'. Overall, there was a trend for basic residues in alternating subsites and smaller residues in the primed sites compared with the non-primed sites. In addition, there were strict requirements for the amino acids in subsites $S_3 - S_1$. Fluorescence-quenched peptides from the library with the highest on-resin cleavage were resynthesised and their kinetics of hydrolysis by CPB2.8ΔCTE assessed in solution phase assays. Several good substrates containing the quintessential dipeptide particular to cathepsin-L-like enzymes, -F-R/K-, in P_2 and P_1 were identified (e.g. Y(NO$_2$)-EK**FR**↓RGK-K(Abz)G, Abz = 2-aminobenzoyl; $k_{cat}K_m^{-1} = 4298\ mM^{-1}s^{-1}$). However, novel substrates containing the dipeptide -L/I-Q- in P_2 and P_1 were also well hydrolysed (e.g. Y(NO$_2$)-Y**LQ**↓GIQK-K(Abz)G; $k_{cat}K_m^{-1} = 2583\ mM^{-1}s^{-1}$). The effect of utilising different fluorescent donor – quencher pairs on the value of $k_{cat}K_m^{-1}$ was examined. Generally, the use of the Abz/Q-EDDnp donor – quencher pair (EDDnp = N-(2,4-dinitrophenyl)ethylenediamine) instead of K(Abz)/Y(NO$_2$) resulted in higher $k_{cat}K_m^{-1}$ values for analogous substrates.*

KEYWORDS:

combinatorial chemistry · enzyme catalysis · fluorescence resonance energy transfer · hydrolases · proteases

Introduction

The protozoal parasites of the genus *Leishmania*, the causative agent of leishmaniasis, express, in a stage-regulated manner, high levels of several classes of cysteine proteases (CPs) belonging to the papain family. In *Leishmania mexicana*, one group of cysteine proteases, CPBs (cysteine proteases of group B),[2] are expressed from a single tandem array comprising 19 gene copies (*CPB1 – CPB19*), with individual genes encoding subtly different isoenzymes.[3] It is thought that the CPs are central to the survival of the parasite in a mammalian host.[4, 5] The precise mechanisms involved are yet to be elucidated, however, one probability entails the internalisation of MHC class II molecules of the host macrophages by the amastigotes followed by their degradation within parasitic lysosomes by the parasite cysteine proteases. This may be a means of circumventing the host's immune system.[6, 7] *CPB* null mutants show reduced infectivity to macrophages in vitro and also to BALB/c mice, compared with wild-type parasites. Re-expression of a single isoform from the CPB array (CPB2.8) restored infectivity towards macrophages to wild-type levels and implicate these CPB isoenzymes as virulence factors.[8] Inhibitors of the CPB isoenzymes have been shown to reduce the infectivity of wild-type parasites both in vitro towards macrophages[8, 9] and in vivo towards BALB/c mice,[10] thus providing further evidence that these CPB isoenzymes are virulence factors.

Mature cysteine proteases of this class from *L. mexicana* are composed of two domains: a catalytic domain of 219 amino acids and a C-terminal extension (CTE) typically 100 amino acids

[a] *Prof. M. Meldal, Dr. P. M. St. Hilaire*
Carlsberg Laboratory, Department of Chemistry
Gamle Carlsberg Vej 10, 2500 Valby (Denmark)
Fax: (+45) 3327-4708
E-mail: mpm@crc.dk

[b] *L. C. Alves, Prof. M. A. Juliano, Prof. L. Juliano*
Department of Biophysics, Escola Paulista de Medicina
Universidade Federal de São Paulo
Rua Tres de Maio 100, 04044-20 São Paulo (Brazil)

[c] *Dr. S. J. Sanderson, Prof. G. H. Coombs*
Division of Infection and Immunity, Joseph Black Building
University of Glasgow, Glasgow G12 8QQ (UK)

[d] *Dr. J. C. Mottram*
Wellcome Centre for Molecular Parasitology, Anderson College
University of Glasgow, Glasgow G11 6NU (UK)

long.[3, 11] The function of the CTE has long been debated and their role in the targeting of the protein to lysosomes has been theorised. However, the evidence so far points to the fact that a full-length CTE is not required for enzyme trafficking and is not essential for the catalytic activity of the enzyme.[3, 12, 13] Recently, the prodomain has been shown to be sufficient for trafficking of the cysteine proteases to lysosomes.[14]

Because of their tremendous importance in the etiology of leishmaniasis, the parasite's CPs are attractive alternative targets for therapeutic agents that ameliorate the symptoms of leishmaniases. The currently most viable chemotherapeutic treatment of the disease is unsatisfactory due to high toxicity[15] and the development of drug-resistant parasites.[16, 17] As a preface to the development of specific inhibitors of the cysteine proteases of *L. mexicana*, we first characterised the substrate specificity of the CPB2.8 isoform (implicated in the restoration of infectivity towards macrophages[8]) by screening an intramolecularly fluorescence-quenched peptide library.[18–21] A characterisation of the substrate specificity is fundamental for understanding the subtle differences between the isoforms of the enzyme, for establishing that recombinant enzymes lacking the CTE have the same activity and specificity as the native enzymes, for designing inhibitors and for screening of inhibitor libraries as previously demonstrated.[22, 23] We present herein the first systematic characterisation of the substrate specificity of cysteine protease CPB2.8 from *L. mexicana*[3, 24] by the use of a fluorescence-quenched combinatorial peptide library.

Editorial Advisory Board Member:[*]
Morten Meldal,
was born in Denmark in 1954 and is a Senior Scientist heading the Synthesis Group at Carlsberg Laboratory and also an Adjunct Professor at the Danish Technical University. He directs the combinatorial chemistry research at the SPOCC Center, hosted at the Carlsberg Laboratory. He received the 1990 Danish Chemistry Award, the 1996 NKT Research Award for Chemistry, the 1996 Leonidas Zervas Award, and the 1997 Ellen and Niels Bjerrum Gold Medal in Chemistry. He has over 190 publications and reviews in international journals and has filed 10 patents. His research interests include: solid-phase organic combinatorial chemistry, polymer chemistry, peptide and glycopeptide synthesis, glycobiology and molecular immunology, methodology and automation, and enzymology and enzyme inhibition. He is well-known for the development of novel PEG-based supports suitable for both solid-phase synthesis and screening, and for novel solid-phase assays based on fluorescence resonance energy transfer (FRET) which are used for the determination of enzyme activity and inhibition.

[*] *Members of the Editorial Advisory Board will be introduced to the readers with their first manuscript.*

Results

Library synthesis

The fluorescence-quenched library of the general structure shown in Figure 1 was synthesised on 1.7 g $PEGA_{4000}$** resin (200 – 800 μm beads).[20] Seven positions ($X_1 - X_7$) were randomised by using all 20 genetically encoded amino acids. The peptide sequences were flanked by the fluorescence donor AbzOH attached to the side chain of lysine (→K(Abz)), and by the

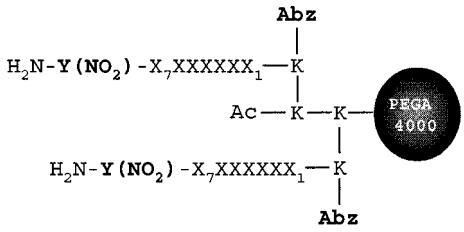

Figure 1. *Peptide library construct. The library, which contained free N termini, was synthesised on $PEGA_{4000}$ resin, and the loading was doubled by incorporation of the sequence of two Lys residues. $X_1 - X_7$ are randomised positions.*

fluorescence quencher, 3-nitrotyrosine (Y(NO_2)). The Lys-Lys sequence was incorporated to double the loading of the resin; initial loading was 0.13 mmol g^{-1} and final loading was 0.21 mmol g^{-1}. The peptides were synthesised by using the Fmoc/OPfp ester methodology[25] except for the incorporation of K(Abz) and Y(NO_2), which were incorporated by using TBTU/NEM activation.[26] Based on the results of Edman degradation, synthesis of the library proceeded smoothly to afford a single compound per bead.

Screening of peptide libraries

After incubation of the library with CPB2.8ΔCTE, beads of varying degrees of fluorescence were isolated from the library and the amino acid sequence of 41 beads was determined by Edman degradation. In three cases, it was impossible to determine the sequence from the Edman degradation data, therefore, the sequences of 38 beads only are shown in Table 1. The results from the sequencing showed that roughly 42% of the peptides were cleaved at two positions and the alignments resulting from both cleavages are shown (Table 1) and used for the generation of the amino acid subsite frequency diagrams shown in Figure 2. While most of the peptides were roughly centrally cleaved by the enzyme, in about half of the peptides with the N-terminal sequence Y′-R/K/H-X- (Table 1, peptides **1** – **13**) were cleaved after the basic amino acid to release the N-terminal dipeptide in addition to hydrolysis at a primary cleavage site. The percentage of peptide hydrolysed by the enzyme on each bead varied widely.

The results from screening of the library indicated that the hydrolysed peptides were highly basic, with 92% of the sequences containing 2 to 4 and 49% containing 3 to 4 basic residues. The specificity requirements for the non-primed site

[**] For abbreviations see ref. [1].

Table 1. *Peptide sequences obtained from screening of the peptide library (Y′-$X_7X_6X_5X_4X_3X_2X_1$-K′) with CPBΔCTE.[a, b]*

Compd.	Peptide Sequence														Hydr. [%][c]
	P_7	P_6	P_5	P_4	P_3	P_2	P_1	↓	P_1'	P_2'	P_3'	P_4'	P_5'	P_6'	P_7'	
1a						Y′	?	↓	E	Q	E	R	K	M	K′	n.d.
2a						Y′	R	↓	F	–	Y	–	S	–	K′	21
3a						Y′	R	↓	A	R	S	G	K	A	K′	18
4a						Y′	R	↓	S	T	I	R	R	G	K′	25
5a						Y′	R	↓	N	S	N	A	N	R	K′	25
6						Y′	R	↓	N	Y	N	N	F	R	K′	16
7a						Y′	K	↓	T	W	R	P	V	R	K′	11
8a						Y′	K	↓	K	I	R	Y	A	D	K′	16
9						Y′	K	↓	K	G	R	V	P	Y	K′	64
10						Y′	K	↓	L	H	P	K	L	K	K′	26
11a						Y′	L	↓	Q	H	L	H	H	N	K′	10
12a						Y′	H	↓	R	A	W	N	V	R	K′	13
13a						Y′	Q	↓	N	K	R	S	R	M	K′	16
14a					Y′	R	I	↓	I	R	I	K	R	K′		n.d.
15a					Y′	I	R	↓	W	R	Q	Q	Y	K′		6
16a					Y′	I	R	↓	R	Y	L	R	E	K′		4
17					Y′	L	K	↓	A	P	Y	S	R	K′		65
18					Y′	L	K	↓	A	K	K	M	G	K′		37
19a					Y′	L	K	↓	L	L	R	K	M	K′		12
20					Y′	V	K	↓	T	P	K	T	S	K′		8
21				Y′	Q	L	V	↓	L	Q	C?	V	K′			31 – 93
22				Y′	Y	L	Q	↓	G	I	Q	K	K′			38 – 45
23				Y′	K	I	Q	↓	V	I	K	G	K′			75
24				Y′	K	L	F	↓	N	P	K	F	K′			70
25				Y′	R	F	F	↓	R	N	R	F	K′			54
26a				Y′	T	V	K	↓	Y	K	V	P	K′			73
27[d]				Y′	R	I	K	↓	R/S	N	I	S	K′			55 – 62
28				Y′	M	F	K	↓	G	I	W	K	K′			54
2b				Y′	R	F	–	↓	Y	–	S	–	K′			64
3b				Y′	R	A	R	↓	S	G	K	A	K′			51
29a				Y′	K	L	S	↓	K	Y	L	S	K′			36
4b				Y′	R	S	T	↓	I	R	R	G	K′			9
7b				Y′	K	T	W	↓	R	P	V	R	K′			9
30			Y′	R	T	L	K	↓	A	R	R	K′				80
31			Y′	K	?	I	K	↓	S	K	K	K′				n.d.
32			Y′	P	K	F	R	↓	S	F	N	K′				67 – 76
33			Y′	E	K	F	R	↓	R	G	K	K′				70
34			Y′	P	R	F	R	↓	T	G	S	K′				45 – 70
8b			Y′	K	K	I	R	↓	Y	A	D	K′				68
15b			Y′	I	R	W	R	↓	Q	Q	Y	K′				5
13b			Y′	Q	N	K	R	↓	S	R	M	K′				29
35			Y′	P	Q	A	M	↓	A	F	R	K′				32 – 42
26b		Y′	T	V	K	Y	K	↓	V	P	K′					7 – 17
36		Y′	I	E	R	F	K	↓	M	R	K′					60
11b		Y′	L	Q	H	L	H	↓	H	N	K′					60
37		Y′	P	Y	R	F	H	↓	T	P	K′					51
2b		Y′	R	F	-	Y	-	↓	S	-	K′					26
4c		Y′	R	S	T	I	R	↓	R	G	K′					28
19b		Y′	L	K	L	L	R	↓	K	M	K′					7 – 8
1b		Y′	?	E	Q	E	R	↓	K	M	K′					95
29b	Y′	K	L	S	K	Y	L	↓	S	K′						33
12b	Y′	H	R	A	W	N	V	↓	R	K′						25
5b	Y′	R	N	S	N	A	N	↓	R	K′						42
16b	Y′	I	R	R	Y	L	R	↓	E	K′						37
38	Y′	A	N	H	F	L	R	↓	T	K′						13 – 30
3c	Y′	R	A	R	S	G	K	↓	A	K′						16
14b	Y′	R	I	I	R	I	K	↓	R	K′						21

[a] Assay conditions: 0.1 M Na phosphate, 137 mM NaCl, 10 mm DTT, pH 6.0, 37 °C. [b] ↓ = cleavage site. Peptides with more than one cleavage site are designated **a** – **c**. Y′ = Tyr(NO_2), K′ = Lys(Abz). [c] Approximate degree of hydrolysis of resin-bound peptide based on results of Edman degradation. [d] It was not possible to conclusively identify the amino acid as Ser or Arg. n.d. = not determined.

were stricter than for the primed sites, particularly for $S_3 - S_1$. Of the primed sites, S_3' exhibited the strictest requirement. There appeared to be no specificity in the S_4 site with acidic (Glu), basic (Arg and Lys), secondary (Pro) and small polar (Ser) amino acids appearing with equal frequency. There was a preference for Arg or Lys in the S_3 subsite and for hydrophobic residues, both

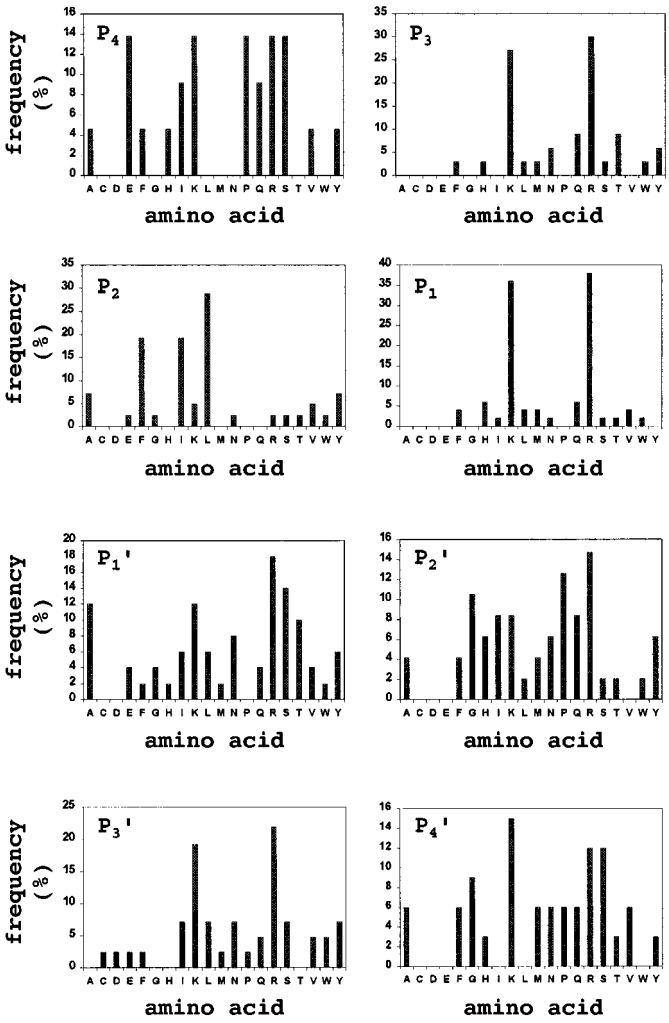

Figure 2. *Amino acid frequency in enzyme subsites. Peptide sequences were identified from screening of the fluorescence-quenched peptide library and are shown in Table 1.*

aromatic and aliphatic, in S_2. In that position, Phe and Ile occurred with equal frequency and were slightly surpassed by Leu. The S_1 subsite exhibited a specificity for the basic residues Arg and Lys. The S_1' site showed favour for Arg, Ser, Lys and Ala, while Arg, Pro and Gly were slightly predominant in S_2'. As for S_1 and S_3, a strict preference for the basic residues Arg and Lys was once more observed in S_3'. Overall, there was a trend for smaller residues in the primed sites compared to the non-primed sites and basic residues in the alternating subsites, P_3, P_1 and P_1' and P_3' (thus the general structure was X-R/K-X-R/K↓-R/K-X-R/K). In the event that a basic residue was not present in the expected subsite, it was placed in an adjacent subsite (e.g. R/K-X-X-R/K↓-X-R/K-X and/or X-R/K-X-R/K↓-R/K-X-X-R/K).

Kinetic characterisation of selected peptides from the library

Fluorescence-quenched peptides with the highest on-resin cleavage and primarily one cleavage site were selected from the library and resynthesised. Two different fluorescence donor–quencher pairs, K(Abz)/Y(NO_2) and Abz/Q-EDDnp, were used to evaluate their respective effects on the kinetics of hydrolysis of the selected peptides. The hydrolysis of Abz/Q-EDDnp peptides was performed at pH 5.5, at which the enzyme exhibits maximal activity (Figure 3). Hydrolysis of K(Abz)/Y(NO_2) peptides was carried out at pH 6.5, at which the enzyme still retains about 95% of maximal activity and is longer active than at pH values below 6.0. These experiments were carried out at higher pH than for Abz/Q-EDDnp peptides in order to optimise the quenching effect of Y(NO_2), although K(Abz)/Y(NO_2) substrates have been useful at low pH.[27]

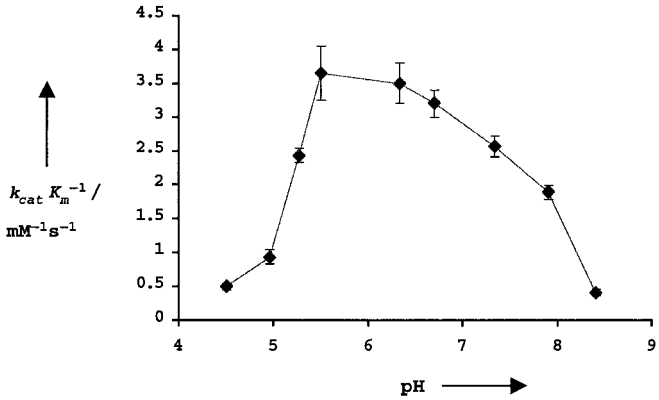

Figure 3. *pH profile of CPB2.8ΔCTE activity determined at 37 °C, using Abz-EKFRRGKQ-EDDnp as substrate.*

The kinetic parameters for substrate hydrolysis are shown in Table 2, wherein the substrates were sorted by the polarity index of the amino acids at the three most restrictive subsites, P_3, P_2 and P_1. It is clear from the data in Table 2 that the kinetic constants are affected by the use of the different pairs of reporter groups. Frequently, the use of the Abz/Q-EDDnp donor-quenched pair resulted in higher $k_{cat}K_m^{-1}$ values for the analogous substrates due either to faster rates of hydrolysis, lower K_m values or a combination of both. In a few instances, the effect of changing reporter groups was quite dramatic, making good substrates resistant to hydrolysis (see **18**, **24** and **30** in Table 2). In two instances, the cleavage points of the substrates were different, as in peptides **8** and **11** (Table 2) where the Y(NO_2) residue fits to the S_2 subsite and the enzyme acts as an amino-dipeptidase. The effects of the fluorescence donor and quencher groups depend on both the peptide sequence and on their distance from the scissile bond, as was previously reported for papain.[21] These variations in kinetic constants notwithstanding, the same general trends of the CPB2.8ΔCTE specificity towards the identified peptide sequences were observed for both series of substrates.

The substrates with the highest specificity constant had the preferred amino acid in most restrictive subsites (P_3–P_1) and generally contained the basic-hydrophobic-basic motif in those sites (Table 2). They also had the highest on-resin cleavage (i.e. 60–80%, Table 1). These substrates had the highest k_{cat} values and bound less tightly to the enzyme (higher K_m). The presence of Phe in P_2 seemed particularly advantageous compared to Ile or Leu. In some cases, the lack of a basic residue in P_3 was compensated by having it in P_4 (peptides **16** and **30**). Substrates

Table 2. Kinetic parameters for CPB2.8ΔCTE hydrolysis of selected peptides identified from the library screen.[a]

Compd.	Sequence	K_m [μM]	k_{cat} [s^{-1}]	$k_{cat}K_m^{-1}$ [$mM^{-1}s^{-1}$]	K_i [μM]
P_3 (basic)-P_2 (hydrophobic)-P_1 (basic)					
8a,b	Y′-K ↓ **KIR** ⇓ YAD-K′G	1.34	0.60	462	n.d.
	Abz-K**KIR** ⇓ YAD-Q′	1.40	1.12	784	n.d.
11a,b	Y′-L ↓ Q**HLH** ⇓ HN-K′G	0.71	0.55	775	n.d.
	Abz-LQ ⇓ **HLH**HN-Q′	0.20	1.15	5793	n.d.
26a,b	Y′-TV**K** ↓ **YK** ⇓ VP-K′G	0.64	0.89	1390	n.d.
	Abz-TV**K** ↓ **YK** ⇓ VP-Q′	0.47	0.46	978	n.d.
27	Y′-**RIK** ⇓ RNIS-K′G	0.50	0.43	877	n.d.
	Abz-**RIK** ⇓ RNIS-Q′	0.30	0.47	1538	n.d.
32	Y′-P**KFR** ⇓ SFN-K′G	0.15	0.62	4133	n.d.
	Abz-YP**KFR** ⇓ SFN-Q′	0.40	2.75	6906	n.d.
33	Y′-E**KFR** ⇓ RGK-K′G	0.57	2.45	4298	n.d.
	Abz-E**KFR** ⇓ RGK-Q′	3.0	7.8	2700	n.d.
34	Y′-P**RFR** ⇓ TGS-K′G	1.13	1.47	1300	n.d.
	Abz-YP**RFR** ⇓ TGS-Q′	0.77	1.85	2385	n.d.
36	Y′-IE**RFK** ⇓ MR-K′G	0.33	1.14	3413	n.d.
37	Y′-PY**RFH** ⇓ TP-K′G	1.80	1.40	778	n.d.
	Abz-YPY**RFH** ⇓ TP-Q′	1.70	1.41	817	n.d.
P_3 (hydrophobic)-P_2 (hydrophobic)-P_1 (basic)					
16a,b	Abz-IR ↓ R**YLR** ⇓ E-Q′	0.36	0.80	2300	n.d.
17	**Y′-LK** ⇓ APYSR-K′G	0.31	0.51	1645	n.d.
	YLKAKKMG-Q-NH_2	n.d.	n.d.	n.d.	0.330[b]
18	**Y′-LK** ⇓ AKKMG-K′G	0.39	0.77	1974	n.d.
	Abz-**YLK**AKKMG-Q′	n.d.	n.d.	n.d.	0.018
25	Abz-YR**FFR**NRF-Q′	n.d.	n.d.	n.d.	0.039
28	Y′-**MFK** ⇓ GIWK-K′G	0.34	0.31	912	n.d.
	Abz-**MFK** ⇓ GIW K-Q′	0.10	0.78	7590	n.d.
P_3 (basic)-P_2 (hydrophobic)-P_1 (hydrophobic)					
14a,b	Abz-**RII**RIKR-Q′	n.d.	n.d.	n.d.	0.042
19a,b	Abz-L**KLL**RKM-Q′	n.d.	n.d.	n.d.	0.012
24	Y′-**KLF**NPKF-K′G	n.d.	n.d.	n.d.	0.031
	Abz-**KLF** ⇓ NPKF-Q′	0.17	0.70	4118	n.d.
29a,b	Y′-KLS**KYL**S-K′G	n.d.	n.d.	n.d.	0.036
	Abz-KLS**KYL**SK-Q′	n.d.	n.d.	n.d.	0.007
38	Abz-AN**HFL** ⇓ CT-Q′	0.44	1.47	3344	n.d.
other combinations of P_3, P_2 and P_1					
21	Y′-**QLV** ⇓ LQCV-K′G	0.56	0.14	259	n.d.
22	Y′-**YLQ** ⇓ GIQK-K′G	0.24	0.62	2583	n.d.
	Abz-**YLQ** ⇓ GIQK-Q′	0.20	0.67	3300	n.d.
23	Y′-**KIQ** ⇓ VIKG-K′G	0.18	0.40	2555	n.d.
	Abz-**KIQ** ⇓ VIKG-Q′	0.66	0.60	913	n.d.
30	Y′-R**TLK**ARR-K′G	n.d.	n.d.	n.d.	0.075
	Abz-R**TLK** ⇓ ARR-Q′	0.14	1.10	7597	n.d.

[a] ⇓ = major cleavage site; ↓ = minor cleavage site. For peptides **8**, **11** and **26**, the ratio of major:minor product was 70:30, for peptide **16**, it was 90:10. [b] This peptide was hydrolysed with a velocity similar to that of its analogue containing the Y(NO_2)/K(Abz) pair. n.d. = not determined.

having the hydrophobic-hydrophobic-basic sequence in the P_3 to P_1 subsites had fairly high $k_{cat}K_m^{-1}$ values primarily due to tighter binding (lower K_m) than those with the basic-hydrophobic-basic motif. To a lesser extent, substrates with the -L/I-Q motif in P_2 and P_1 were hydrolysed well. In agreement with the library screening results, the requirements in the primed subsites were less strict, with most substrates containing one or more of the preferred residues in positions $P_1'-P_3'$.

Some of the peptides presented in Table 2 (**14**, **19**, **24**, **29** and **38**) are particularly interesting, since they contain the basic-hydrophobic-hydrophobic sequence of amino acids in P_3-P_1 and were slowly hydrolysed by the enzyme both in solution phase assays as well as on solid phase during the library screening (Table 1). The affinity of these peptides for CPB2.8ΔCTE was evaluated by assaying them as inhibitors, and the K_i values were obtained from typical competitive inhibition reactions using Z-Phe-Arg-AMC as the substrate. Exceptions to this general trend are peptides **24** and **38**, both labelled with Abz/Q-EDDnp, which were efficiently hydrolysed by the enzyme.

Discussion

Cysteine proteases from trypanosomatid parasites are important targets for a new generation of therapeutic agents. It is crucial that these protease inhibitors exhibit high specificity for the parasite's enzyme. Since the genes encoding many of these enzymes are in tandem arrays resulting in isoforms with subtly different substrate specificities,[3] it is important for the design of highly specific inhibitors that the substrate specificity of these enzymes be characterised. To date, there has been limited systematic characterisation of the entire substrate specificity of protozoal proteases. Their specificity has generally been characterised by conventional methods involving the systematic yet tedious synthesis of several substrates usually containing a chromophore (AMC or pNA) at one end. This approach is limited because it is practically impossible to synthesise all the possible different substrates for testing and, furthermore, the use of chromogenic substrates gives rise to information about the requirements of the non-primed (S) or primed subsites (S′) independently. In addition, it has been demonstrated that subsite–substrate interaction is generally dependent on the substrate structure and is not necessarily additive.[28] The use of an internally quenched fluorescent substrate,[27, 29, 30] on the other hand, gives information about both the non-primed and primed subsites simultaneously, making the use of a fluorescence-quenched peptide library (in which case the enzyme itself selects substrates for hydrolysis) ideal for the investigation of the substrate specificity of proteolytic enzymes.[18, 19, 21, 31]

The screening of a fluorescence-quenched combinatorial peptide library has resulted in the characterisation of the substrate specificity in both the non-primed and primed subsites of the CPB2.8 isoform (lacking the C-terminal extension) from *L. mexicana*. The specificity of the enzyme determined herein is similar to that of the related enzyme from *Trypanosoma cruzi*, cruzipain,[19] with which it shares a 59% homology in the mature domain.[11] In CPB2.8ΔCTE, the most restrictive subsites are S_3, S_2 and S_1, and the best substrates (e.g. Abz-YP**KFR** ↓ SFN-Q′ (**32**) and Y′-E**KFR** ↓ RGK-K′G (**33**)) share the basic-hydrophobic-basic motif in the P_3-P_1 positions. By contrast, peptides containing the basic-hydrophobic-hydrophobic sequence of amino acids in these positions were slowly hydrolysed and inhibited the enzyme. The inhibitory nature of these peptides may be related to their resemblance to the **RYL**NGA sequence of the proregion of CPB2.8.[3] It is well documented that cathepsins B and L and cruzipain are inhibited by their respective proregions.[32, 33]

The specificity of the primed sites of CPB2.8ΔCTE was less strict than that of the non-primed sites and favoured small amino acids. However, there was at least one basic residue in $P_3'-P_1'$. Generally, there were basic residues in the alternating

subsites, P_3 and P_1, and P_1' and P_3', and in the event that a basic residue was not present in the expected subsite, it was placed in an adjacent subsite. This preference of basic residues in alternating subsites might be correlated to the nature of the amino acids in the active site. It is possible, as in the case of cruzain, that basic residues in the substrates form salt bridges with some of the acidic residues that line the active site such as Asp 18, Glu 112, Asp 158, Asp 64, Asp 61 and Glu 205 (papain numbering).[11, 34] In addition to the expected substrates described above, a new type of substrate containing Ile or Leu in P_2 and Gln in P_1 was identified by using the library method: Abz-YLQ↓GlQK-Q′ (**22**) and Y′-KIQ↓VIKG-K′G (**23**). This finding is in agreement with results of experiments with cruzipain where it was also found that substrates with Leu in P_2 were well hydrolysed by the enzyme.[19] The acceptance of Gln in P_1 is probably related to the fact that Gln can interact with a carboxylate group through hydrogen bonding in a similar manner as the guanidino group of Arg[35] and is one of the preferred structural replacements for exposed arginines in protein mutations.[36]

A close examination of the good substrates possessing the basic-hydrophobic-basic motif in P_3-P_1 reveals that Phe most frequently occupies the P_2 site. This observation is in agreement with previous experiments using short AMC substrates that demonstrated the cathepsin-L-like specificity of the enzyme.[24, 37] Substrates with a hydrophobic residue (Phe) in P_2 and Arg in P_1 were hydrolysed to a larger extent than those with Arg in P_2. The cathepsin-B-like enzyme from *Leishmania major* which has an 80% sequence homology to CPB2.8 exhibits a similar specificity towards small substrates.[38, 39] It is noteworthy that of the 57 cleaved sequences obtained from the library screen, only two sequences (**13b** and **14a**) were hydrolysed when a basic residue occupied P_2.

An analysis of the cleaved sequences shown in Table 2 suggests that the CPB2.8ΔCTE enzyme can act as an amino-dipeptidase. Substrates containing the N-terminal sequence Y′-R/K/H-X- (Table 1, peptides **1** – **13**) were cleaved after the basic amino acid to release the N-terminal dipeptide in addition to hydrolysis at a primary cleavage site. This ability to accommodate $Y(NO_2)$ at S_2 is in keeping with recent observations that the presence of electronegative substituents on an aromatic ring at P_2 enhances the interaction of the substrate with the enzyme.[21, 40] Another interesting finding from the library results is that Pro is a favoured residue in P_2' and substrates containing Pro in that position are fairly well cleaved by the enzyme (Table 2, peptides **17**, **26** and **37**). This finding is also true for cruzipain[19] and may reflect the substrate specificity of the enzyme during auto-processing to the mature, active form (cleavage sites for CPB2.8: proregion: --DLSA↓VPDA--, CTE: --RESA↓APGT--).

It must be noted that while the use of different donor–quencher groups resulted in different kinetic constants, the specificity requirement in the respective enzyme subsites was generally maintained. However, in a few instances, simply changing the nature of the reporter groups induced a radical conversion from substrate to inhibitor (and vice versa). This finding underscores the interdependency between an enzyme's subsites and the fact that subtle changes at the distal ends of a substrate can affect its entire orientation or conformation, making catalysis more facile or impossible. As observed from modelling studies with papain, small changes in the substrate (e.g. changing one amino acid) can cause its alignment in the active site to shift such that the sulfur atom of the cysteine residue is no longer able to attack the carbonyl C atom of the peptide bond and attacks an adjacent carbonyl group instead.[21] A more detailed structural analysis of this conversion from substrate to inhibitor (and vice versa) will be carried out once a crystal structure of the enzyme is obtained.

In conclusion, the screening of fluorescence-quenched substrate library with isoform CPB2.8 from *Leishmania mexicana* has allowed the enzyme to select several good substrates, both those expected because of its similarity to cruzipain as well as other unexpected, but logical ones.

Experimental Section

General: Dichloromethane and DMF were dried, distilled, and stored over 3 Å molecular sieves under inert gas. Fmoc amino acids and their Pfp ester derivatives were purchased from Bachem and NovaBiochem. Fmoc-Lys(Boc-Abz)-OH, Fmoc-Tyr(NO_2)-OH and Fmoc-Glu-EDDnp were prepared as previously described.[27, 41] The substitution of the resins was determined by spectrophotometric analysis (Perkin–Elmer Lambda 7 UV/Vis spectrophotometer) at 290 nm of the dibenzofulvenepiperidine adduct formed upon deprotection of the amino-terminal Fmoc group. Purification of Y′-peptidyl-K′ peptides was performed by preparative reverse-phase HPLC on a Waters HPLC system with a delta pak C-18 column (200 × 25 mm) and a linear gradient of solvent A (0.1% TFA in water) and solvent B (0.1% TFA in 90% aq. MeCN) at a flow rate of 20 mL min^{-1}. Amino acid sequencing was performed on a Shimadzu Sequencer model PPSQ-23 for peptides in solution, and on Applied Biosystems Sequencer models 477A or 470A equipped with an on-line phenylthiohydantoin analyzer (Model 120A) for resin-bound substrates. MALDI-TOF MS of synthetic peptides, Y′-peptidyl-K′, was performed on a Finnigan Lasermat 2000 with a matrix of α-cyano-4-hydroxycinnamic acid. The Abz-peptidyl-Q-EDDnp substrates and their fragments were analysed in the reflectron mode on a TofSpec-E instrument from Micromass.

Enzyme expression and purification: *L. mexicana* CPB2.8ΔCTE was expressed, purified and activated as previously described.[24] The concentration of the enzyme stock solution (11.4 μM) was determined by active-site titration with human cystatin (a generous gift from Dr. Magnus Abrahamson, University of Lund, Sweden) using Z-Phe-Arg-AMC as the substrate.

Determination of the pH profile of the enzyme: The activity of CPB2.8ΔCTE was determined over the pH range 4.5 to 8.5 at 37 °C, using Abz-EKFRRGKQ-EDDnp as substrate and 0.74 nM enzyme. The following buffers augmented with 2 mM EDTA and 200 mM NaCl were used: 0.1 M sodium acetate for pH 4.5 to 5.5, 0.1 M citrate/phosphate for pH 5.5 to 6.5, 0.1 M sodium phosphate for pH 6.5 to 7.5 and 0.1 M tris/HCl for pH 7.5 to 8.5.

General methods for solid-phase peptide synthesis: Synthesis of the peptide substrates and the library of general structure $Y'X_7X_6X_5X_4X_3X_2X_1K'$ was carried out manually by MCPS[42, 43] on PEGA resin in a 20-column teflon synthesis block as previously described.[21] Reaction completion was assessed by using either Dhbt-OH (1 equiv), which served both as an acylation catalyst and an indicator of the reaction completion,[44] or by the Kaiser test.[45]

Solid-phase substrate library synthesis: The library, $Y(NO_2)X_7X_6X_5X_4X_3X_2X_1K(Abz)$ (Figure 1), containing all 20 genetically encoded amino acids was prepared on $PEGA_{4000}$ resin as previously described.[21, 46]

Solid-phase multiple-column synthesis of Y′-peptidyl-K′: Lead peptides from the library screen were synthesised on $PEGA_{800}$ resin (150 – 300 μm beads, loading = 0.34 mmol g^{-1}, 19 μM peptide per well) as previously described.[21] After cleavage from the resin, the peptides were characterised by MALDI-TOF MS and their purity assessed by analytical HPLC.

Solid-phase synthesis of Abz-peptidyl-Q-EDDnp: Peptides were synthesised according to the manufacturers protocol by using the Fmoc methodology on an automated bench-top simultaneous multiple solid-phase peptide synthesiser (PSSM 8 system from Shimadzu) as previously described.[41] After deprotection, the peptides were purified by semi-preparative HPLC using an Econosil C-18 column (10 μ, 22.5 × 250 mm) and a gradient of solvent A and solvent B. The column was eluted at a flow rate of 5 mL min^{-1} with a 10 → 50 (or 30 → 60) % gradient of solvent B over 30 (or 45) min. Analytical HPLC was performed on a binary HPLC system from Shimadzu which was equipped with a SPD-10AV Shimadzu UV-Vis detector and a Shimadzu RF-535 fluorescence detector, coupled to an Ultrasphere C-18 column (5μ, 4.6 × 150 mm). Peptides were eluted with solvent systems A1 (H_3PO_4/H_2O, 1:1000) and B1 (MeCN/H_2O/H_3PO_4, 900:100:1) at a flow rate of 1.7 mL min^{-1} and a 10 → 80 % gradient of B1 over 15 min. The HPLC column eluates were monitored for absorbance at 220 nm and fluorescence emission at 420 nm upon excitation at 320 nm. The masses of the synthesised peptides were analysed by MALDI-TOF MS.

Solid-phase library screening: The library (300 mg, ca. 50 000 beads) was washed (2 × 10 min) with assay buffer (100 mM phosphate buffer, pH 6.0, augmented with 10 mM DTT) then incubated with activated CPB2.8ΔCTE (100 μL of stock solution; 117 nM) at 37 °C in 7 mL assay buffer. The fluorescence intensity of the beads was monitored with a fluorescence microscope every 30 min for indications of hydrolysis. After 3 h, several beads showed a fluorescent "ring" indicating hydrolysis of some of the peptides on the beads. The reaction mixture was treated with cysteine protease inhibitor E-64 for 1 h before treatment with 2% aq. TFA solution (2 × 5 min), water (2 ×), 2% $NaHCO_3$ (2 ×), then water (3 ×). The fluorescence intensity of the beads was assessed by inspection with a fluorescence microscope, and bright beads were collected and transferred to a TFA-treated cartridge filter for on-resin sequence analysis. The amino acid sequence and the cleavage point of the peptide substrates were determined by Edman degradation. The extent of cleavage was determined by a comparison of the amount (picomoles) of an amino acid in both the non-cleaved and cleaved peptide in different cycles of the degradation.

Hydrolysis of fluorescence-quenched substrates in solution: Hydrolysis of fluorogenic peptide substrates was carried out using 0.1 M sodium acetate, pH 5.5, augmented with 2 mM EDTA, 200 mM NaCl and 10 mM DTT, for Abz-peptidyl-Q-EDDnp peptides, and 0.1 M sodium phosphate, pH 6.5, augmented with 2 mM EDTA, 200 mM NaCl, and 10 mM DTT for $Tyr(NO_2)$-peptidyl-Lys(Abz)-Gly-OH peptides. All experiments were carried out at 37 °C with enzyme pre-incubation in 10 mM DTT for 15 min at room temperature. Substrate hydrolysis was monitored by measuring the fluorescence emission as previously described.[29, 47] The enzyme concentration varied from 0.006 nM for the best substrates to 12 nM for the less susceptible ones while the substrate concentration ranged from 0.1 K_m to 10 K_m. The kinetic parameters were calculated as previously described[48] and in all cases the standard errors of K_m and k_{cat} determinations were less than 5 % of the obtained value. When peptides were either resistant to hydrolysis, or were slowly hydrolysed (less than 1.5 nmol min^{-1}) at the upper limit of enzyme concentration or inhibited the enzyme at concentrations lower than the estimated K_m value, K_i values for competitive inhibition were determined according to the method of Nicklin and Barrett[49] using Z-Phe-Arg-AMC as the substrate. The cleavage site was determined by correlation of the retention times of the isolated fluorescent Abz-containing peptide fragments with known synthesised sequences. Alternatively, amino acid sequencing or fragment mass determination by MALDI-TOF MS was used.

This work was supported by the INCO-DC program (EU Contract no. ERBIC18CT970225), and the Brazilian Research Foundations FAPESP and PADCT. We acknowledge assistance from: Dr. Ib Svendsen, Bodil Corneliussen and Lone Sørensen for peptide sequencing and Hanne Christiansen for peptide purification (Denmark); Dr. Izaura Y. Hirata for peptide sequencing and Egleliza Galucci de Andrade for peptide synthesis and purification (Brazil); and Kevin Pollock for enzyme production (UK). J.C.M. is an MRC Senior Fellow (UK).

[1] Abbreviations: Abz = 2-aminobenzoyl, AMC = 7-amino-4-methylcoumarin, Boc = *tert*-butoxycarbonyl, CHC = α-cyano-4-hydroxycinnamic acid, CPB2.8ΔCTE = cysteine protease group B isoenzyme lacking C-terminal extension, Dhbt-OH = 3,4-dihydro-3-hydroxy-4-oxo-1,2,3-benzotriazine, DMF = dimethylformamide, DTT = dithiothreitol, EDDnp = *N*-(2,4-dinitrophenyl)ethylenediamine, Fmoc = N^{α}-fluoren-9-ylmethoxycarbonyl, HMBA = hydroxymethylbenzoic acid, MCPS = multiple-column peptide synthesis, MeIm = *N*-methylimidazole, MSNT = 1-(mesitylene-2-sulfonyl)-3-nitro-1*H*-1,2,4-triazole, NEM = *N*-ethylmorpholine, PEGA = polyethylene glycol – acrylamide copolymer, Pfp = pentafluorophenyl, Pmc = 2,2,5,7,8-pentamethylchroman-6-sulfonyl, PNA = *p*-nitroanilide, TBTU = *O*-(benzotriazol-1-yl)-*N*,*N*,*N*′,*N*′-tetramethyluronium tetrafluoroborate, TFA = trifluoroacetic acid, ↓ = minor cleavage site, ⇓ = major cleavage site.

[2] C. D. Robertson, G. H. Coombs, *Microbiology* **1994**, *140*, 417 – 424.

[3] J. C. Mottram, M. J. Frame, D. R. Brooks, L. Tetley, J. E. Hutchison, A. E. Souza, G. H. Coombs, *J. Biol. Chem.* **1997**, *272*, 14 285 – 14 293.

[4] G. H. Coombs, J. C. Mottram, *Parasitology* **1997**, *114*, S61-S80.

[5] G. H. Coombs, J. C. Mottram in *Trypanosomiasis and Leishmaniasis* (Eds.: G. Hide, J. C. Mottram, G. H. Coombs, P. H. Holmes), CAB International, **1997**, pp. 177 – 197.

[6] L. S. DeSouza, T. Lang, E. Prina, R. Hellio, J. C. Antoine, *J. Cell. Sci.* **1995**, *108*, 3219 – 3231.

[7] J. Alexander, G. H. Coombs, J. C. Mottram, *J. Immunol.* **1998**, *161*, 6794 – 6801.

[8] J. C. Mottram, A. E. Souza, J. E. Hutchison, R. Carter, M. J. Frame, G. H. Coombs, *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 6008 – 6013.

[9] M. J. Frame, J. C. Mottram, G. H. Coombs, *Parasitology* **2000**, in press.

[10] P. M. Selzer, S. Pingel, I. Hsieh, B. Ugele, V. Chan, J. C. Engel, M. Bogyo, D. G. Russell, J. Sakanari, J. H. McKerrow, *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 11 015 – 11 022.

[11] A. E. Souza, S. Waugh, G. H. Coombs, J. C. Mottram, *FEBS Lett.* **1992**, *311*, 124 – 127.

[12] A. E. Eakin, M. E. McGrath, J. H. McKerrow, R. J. Fletterick, C. S. Craik, *J. Biol. Chem.* **1992**, *268*, 6115 – 6118.

[13] E. G. Pamer, C. E. Davis, N. So, *Infect. Immun.* **1991**, *59*, 1074 – 1078.

[14] J. A. Huete-Perez, J. C. Engel, L. S. Brinen, J. C. Mottram, J. H. McKerrow, *J. Biol. Chem.* **1999**, *274*, 16 249 – 16 256.

[15] World Health Organization: UNDP/World Bank/WHO 8, Leishmaniasis, Special Programme for Research and Training in Tropical Disease, *Tropical Disease Research: Progress 1991 – 1992*, *Eleventh Programme Report*, **1993**, pp. 77 – 87.

[16] M. Grogl, T. N. Thomason, E. D. Franke, *Am. J. Trop. Med. Hyg.* **1992**, *47*, 117 – 126.

[17] P. Borst, M. Ouellette, *Annu. Rev. Microbiol.* **1995**, *49*, 427 – 460.
[18] M. Meldal, I. Svendsen, K. Breddam, F.-I. Auzanneau, *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 3314 – 3318.
[19] E. Del Nery, M. A. Juliano, M. Meldal, I. Svendsen, J. Scharfstein, A. Walmsley, L. Juliano, *Biochem. J.* **1997**, *323*, 427 – 433.
[20] M. Renil, M. Ferreras, J. M. Delaisse, N. T. Foged, M. Meldal, *J. Peptide Sci.* **1998**, *4*, 195 – 210.
[21] P. M. St. Hilaire, M. Willert, M. A. Juliano, L. Juliano, M. Meldal, *J. Comb. Chem.* **1999**, *1*, 509 – 523.
[22] M. Meldal, I. Svendsen, *J. Chem. Soc. Perkin Trans. 1* **1995**, 1591 – 1596.
[23] M. Meldal, I. Svendsen, L. Juliano, M. A. Juliano, E. Del Nery, J. Scharfstein, *J. Peptide Sci.* **1998**, *4*, 83 – 91.
[24] S. Sanderson, K. Pollock, J. Hilley, M. Meldal, P. M. St. Hilaire, M. A. Juliano, L. Juliano, J. C. Mottram, G. H. Coombs, *Biochem. J.* **2000**, *347*, 383 – 388.
[25] A. Dryland, R. C. Sheppard, *Tetrahedron* **1988**, *44*, 859 – 876.
[26] R. Knorr, A. Trzeciak, W. Bannwarth, D. Gillessen, *Tetrahedron Lett.* **1989**, *30*, 1927 – 1930.
[27] M. Meldal, K. Breddam, *Anal. Biochem.* **1991**, *195*, 141 – 147.
[28] H. Grøn, K. Breddam, *Biochemistry* **1992**, *31*, 8967 – 8971.
[29] J. R. Chagas, L. Juliano, E. S. Prado, *Anal. Biochem.* **1991**, *192*, 419 – 425.
[30] C. Garcia-Echeverria, D. H. Rich, *FEBS Lett.* **1992**, *297*, 100 – 102.
[31] J. C. Spetzler, V. Westphal, J. R. Winther, M. Meldal, *J. Peptide Sci.* **1998**, *4*, 128 – 137.
[32] M. Cygler, J. S. Mort, *Biochimie* **1997**, *79*, 645 – 652.
[33] G. Lalmanach, F. Lecaille, J. R. Chagas, E. Authie, J. Scharfstein, M. A. Juliano, F. Gauthier, *J. Biol. Chem.* **1998**, *273*, 25 112 – 25 116.
[34] D. Turk, G. Guncar, M. Podobnik, B. Turk, *Biol. Chem.* **1998**, *379*, 137 – 147.
[35] D. J. Nicholls, J. Miller, M. D. Scawen, A. R. Clarke, J. J. Holbrool, T. Atkinson, C. R. Goward, *Biochem. Biophys. Res. Commun.* **1992**, *189*, 1057 – 1062.
[36] D. Bordo, P. Argos, *J. Mol. Biol.* **1991**, *217*, 721 – 729.
[37] C. D. Robertson, G. H. Coombs, *Mol. Biochem. Parasitol.* **1993**, *62*, 271 – 280.
[38] P. M. Selzer, X. Chen, V. Chan, M. Cheng, G. L. Kenyon, I. D. Kuntz, J. Sakanari, F. E. Cohen, J. H. McKerrow, *Exp. Parasitol.* **1997**, *87*, 212 – 221.
[39] V. Chan, P. M. Selzer, J. H. McKerrow, J. Sakanari, *Biochem. J.* **1999**, *340*, 113 – 117.
[40] F. Lecaille, C. Serveau, F. Gauthier, G. Lalmanach, *FEBS Lett.* **1999**, *445*, 311 – 314.
[41] I. Y. Hirata, M. H. S. Cezari, C. R. Nakaie, P. Boschcov, A. S. Ito, M. A. Juliano, L. Juliano, *Lett. Peptide Sci.* **1994**, *1*, 299 – 308.
[42] A. Holm, M. Meldal in *Peptides 1988, Proc. Eur. Pept. Symp., 20th ed.* (Eds.: G. Jung, E. Bayer), de Gruyter, Berlin, **1989**, pp. 208 – 210.
[43] M. Meldal, C. B. Holm, G. Bojesen, M. H. Jacobsen, A. Holm, *Int. J. Pept. Protein Res.* **1993**, *41*, 250 – 260.
[44] E. Atherton, L. Cameron, M. Meldal, R. C. Sheppard, *Chem. Commun.* **1986**, 1763 – 1765.
[45] E. Kaiser, R. L. Colescott, C. D. Bossinger, P. I. Cook, *Anal. Biochem.* **1970**, *34*, 595 – 598.
[46] M. K. Christensen, M. Meldal, K. Bock, *J. Chem. Soc. Perkin Trans. 1* **1993**, 1453 – 1460.
[47] M. Meldal, K. Breddam in *Innovation Perspect. SPS* (Ed.: R. Epton), SPCC LTD., Birmingham, **1990**, *1*, 533 – 536.
[48] G. N. Wilkinson, *Biochem. J.* **1961**, *80*, 324 – 332.
[49] M. J. Nicklin, A. J. Barrett, *Biochem. J.* **1984**, *223*, 245 – 253.

Received: April 14, 2000 [F 41]